

# On a Relationship between the Correct Probability of Estimation from Correlated Data and Mutual Information

Yasutada Oohama

**Abstract**—Let  $X, Y$  be two correlated discrete random variables. We consider an estimation of  $X$  from encoded data  $\phi(Y)$  of  $Y$  by some encoder function  $\phi(Y)$ . We derive an inequality describing a relation of the correct probability of estimation and the mutual information between  $X$  and  $\phi(Y)$ . This inequality may be useful for the secure analysis of crypto system when we use the success probability of estimating secret data as a security criterion. It also provides an intuitive meaning of the secrecy exponent in the strong secrecy criterion.

## I. INTRODUCTION

It is well known that the mutual information is a very important quantity for an evaluation of the security of communication system. In the crypto system introduced by Shannon [1] perfect secrecy is defined by the condition that the mutual information between secret data and encrypted data vanishes. In the wiretap channel investigated by Wyner [2] and in the broadcast channel with confidential messages investigated Csiszár and Körner [3], perfect secrecy is defined by an asymptotically vanishing *mutual information rate per channel use* between the secret messages and the channel outputs obtained by the unauthorized user.

In the several recent researches on the information theoretical security, the strong secrecy condition where *the value of mutual information* should asymptotically be zero is well used. Specifically, Hayashi [4] has derived the relevant secrecy exponent function to specify the exponentially decreasing speed (i.e., exponent) of the leaked information under the average secrecy criterion when no cost constraint is considered. Han *et al.*[5] extend his result to the case with cost constraint. The secrecy condition used by Wyner [2] and Csiszár and Körner [3] now called the weak secrecy condition has a clear intuitive meaning that the leak of information rate on the secret messages is asymptotically zero. On the other hand in the strong secrecy criterion the intuitive meaning of the secrecy exponent function does not seem to be so clear.

In this paper we consider a problem which is related to the intuitive meaning the secrecy exponent. Our problem is as follows. Let  $X, Y$  be two correlated discrete random variables. We consider an estimation of  $X$  from encoded data  $\phi(Y)$  of  $Y$  by some encoder function  $\phi(Y)$ . We derive an inequality describing a relation of the correct probability of estimation and the mutual information between  $X$  and  $\phi(Y)$ . This inequality may be useful for the secure analysis

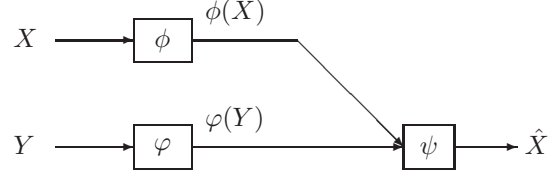


Fig. 1. The case where the side information  $\phi(Y)$  helps an estimation of  $X$  from  $\phi(X)$ .

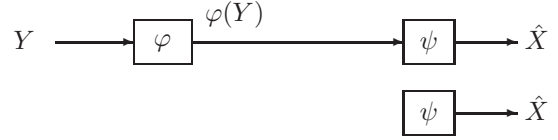


Fig. 2. The case where only one side information is available at the estimator and the case where no information is available at the estimator.

of crypto system when we use the success probability of estimating secret data as a security criterion. It also provides an intuitive meaning of the secrecy exponent in the strong secrecy criterion.

## II. PROBLEM STATEMENT AND RESULTS

### A. Data Estimation from Correlated Data

Let  $\mathcal{X}$  and  $\mathcal{Y}$  be discrete sets. We admit the case where those are countably infinite. Let  $(X, Y)$  be a discrete random pair taking values in  $\mathcal{X} \times \mathcal{Y}$  and having a probability distribution

$$p_{XY} = \{p_{XY}(x, y)\}_{(x, y) \in \mathcal{X} \times \mathcal{Y}}$$

We consider a source estimation system depicted in Fig. 1. Data sequences  $X$  and  $Y$  are separately encoded to  $\phi(X)$  and  $\phi(Y)$  and those are sent to the information processing center. At the center the estimator  $\psi$  observes  $(\phi(X), \phi(Y))$  to output the estimation  $\hat{X}$  of  $X$ . The encoder functions  $\phi$  and  $\varphi$  are defined by

$$\left. \begin{aligned} \phi : \mathcal{X} &\rightarrow \mathcal{M} = \{1, 2, \dots, |\mathcal{M}|\}, \\ \varphi : \mathcal{Y} &\rightarrow \mathcal{L} = \{1, 2, \dots, |\mathcal{L}|\}. \end{aligned} \right\} \quad (1)$$

The estimator  $\psi$  is defined by

$$\psi : \mathcal{M} \times \mathcal{L} \rightarrow \mathcal{X}. \quad (2)$$

Y. Oohama is with University of Electro-Communications, 1-5-1 Chofugaoka Chofu-shi, Tokyo 182-8585, Japan.

The error probability of estimation is

$$P_e(\phi, \varphi, \psi | p_{XY}) = \Pr \left\{ \hat{X} \neq X \right\}, \quad (3)$$

where  $\hat{X} = \psi(\phi(X), \varphi(Y))$ . The correct probability of estimation is

$$P_c(\phi, \varphi, \psi | p_{XY}) = 1 - P_e(\phi, \varphi, \psi | p_{XY}) = \Pr \left\{ \hat{X} = X \right\}. \quad (4)$$

We consider the following three cases.

1. The case where the side information  $\varphi(Y)$  serves as a helper to estimate  $X$  from  $\phi(X)$ . (Case 1)
2. The case where only the helper  $\varphi(Y)$  is available for an estimation of  $X$ . (Case 2) Case 2 corresponds to the case where  $|\mathcal{M}| = 1$  and  $\phi$  is a constant function given by  $\phi(x) = 1, x \in \mathcal{X}$ . The decoder function  $\psi$  in this case is given by  $\psi : \mathcal{L} \rightarrow \mathcal{X}$ .
3. The case where no information is available for an estimation of  $X$ . (Case 3) Case 3 corresponds to the case where  $|\mathcal{M}| = |\mathcal{L}| = 1$  and  $\phi$  and  $\varphi$  are constant functions given by  $\phi(x) = 1, x \in \mathcal{X}$  and  $\varphi(y) = 1, y \in \mathcal{Y}$ . The decoder function  $\psi$  in this case is given by  $\psi : \{1\} \rightarrow \mathcal{X}$ .

Let the correct probability of estimation in Case 2 is denoted by  $P_c(\varphi, \psi | p_{XY})$ . Let the correct probability of estimation in Case 3 is denoted by  $P_c(\psi | p_{XY})$ . Set

$$\begin{aligned} P_{c, \max}^{(1)}(p_{XY}) &\triangleq \max_{\substack{\phi: \mathcal{X} \rightarrow \mathcal{M}, \\ \varphi: \mathcal{Y} \rightarrow \mathcal{L}, \\ \psi: \mathcal{M} \times \mathcal{L} \rightarrow \mathcal{X}}} P_c(\varphi, \psi | p_{XY}), \\ P_{c, \max}^{(2)}(p_{XY}) &\triangleq \max_{\substack{\varphi: \mathcal{Y} \rightarrow \mathcal{L}, \\ \psi: \mathcal{L} \rightarrow \mathcal{X}}} P_c(\varphi, \psi | p_{XY}), \\ P_{c, \max}^{(3)}(p_{XY}) &\triangleq \max_{\psi: \{1\} \rightarrow \mathcal{X}} P_c(\psi | p_{XY}). \end{aligned}$$

Our aim is to clarify relationships between the above three quantities. By definition it is obvious that

$$P_{c, \max}^{(1)}(p_{XY}) \geq P_{c, \max}^{(2)}(p_{XY}) \geq P_{c, \max}^{(3)}(p_{XY}).$$

Set

$$p_{\max} \triangleq \max_{x \in \mathcal{X}} p_X(x).$$

Then we have

$$\begin{aligned} P_{c, \max}^{(3)}(p_{XY}) &= \max_{\substack{\psi: \{1\} \rightarrow \mathcal{X}, \\ x \in \mathcal{X}: \psi(1)=x}} p_X(x) \\ &= \max_{\psi(1) \in \mathcal{X}} p_X(\psi(1)) = p_{\max}. \end{aligned}$$

We are particularly interested in a difference between  $P_{c, \max}^{(2)}(p_{XY})$  and  $P_{c, \max}^{(3)}(p_{XY})$ . If there is no difference between those to quantities. The side information  $\varphi(Y)$  is of no use to estimate  $X$ . In this paper we derive an inequality stating that the difference is upper bounded by the mutual information between the side information  $\varphi(Y)$  and the source  $X$ .

## B. Main Results

In this subsection we state our main result. We first give a proposition which plays a key role in deriving our main results. Set  $S = \varphi(Y)$ . The joint distribution  $p_{XY}$  of  $(X, Y, S)$  is given by

$$p_{XYS}(s, x, y) = p_{XY}(x, y) p_{S|Y}(s|y).$$

It is obvious that the random variables  $X, Y, S$  form Markov chain  $X \leftrightarrow Y \leftrightarrow S$ . The following proposition providing an upper bound of  $P_{c, \max}^{(0)}(p_{XY})$  is useful to derive our main result.

*Proposition 1:* For any  $\eta > 0$  and for any  $(\phi, \varphi, \psi)$ , we have

$$\begin{aligned} P_c(\phi, \varphi, \psi | p_{XY}) \\ \leq p_{SX} \left\{ \log |\mathcal{M}| \geq \log \frac{1}{p_{X|S}(X|S)} - \eta \right\} + 2^{-\eta}. \end{aligned}$$

Specifically, we have

$$\begin{aligned} P_{c, \max}^{(1)}(p_{XY}) \\ \leq p_{SX} \left\{ \log |\mathcal{M}| \geq \log \frac{1}{p_{X|S}(X|S)} - \eta \right\} + 2^{-\eta}. \end{aligned}$$

Proof of this proposition is given in the next section. Using this proposition, we obtain the following result.

*Theorem 1:* For any  $\nu \in (0, \log \frac{1}{p_{\max}})$ , we have

$$P_{c, \max}^{(2)}(p_{XY}) \leq 2^\nu (p_{\max}) + \frac{1}{\nu} I(X; \varphi(Y)). \quad (5)$$

Proof of this theorem is given in the next section. From Theorem 1 and  $2^\nu \leq 1 + \nu$  for  $\nu \in [0, 1]$ , we have the following corollary.

*Corollary 1:* For any  $\nu \in (0, \min\{1, \log \frac{1}{p_{\max}}\})$ , we have

$$P_{c, \max}^{(2)}(p_{XY}) \leq (1 + \nu) p_{\max} + \frac{1}{\nu} I(X; \varphi(Y)).$$

## III. PROOFS OF THE RESULTS

In this section we prove Proposition 1 and Theorem 1. We first prove Proposition 1. To prove this proposition, we prepare a lemma. Set

$$\begin{aligned} \mathcal{D} &\triangleq \{(s, x, y) : s = \varphi(y), p_{X|S}(x|s) \geq (1/|\mathcal{M}|)2^{-\eta}\}, \\ \mathcal{E} &\triangleq \{(s, x, y) : s = \varphi(y), \psi(\phi(x), \varphi(y)) = x\}. \end{aligned}$$

Then we have the following lemma.

*Lemma 1:*

$$p_{SXY}(\mathcal{D}^c \cap \mathcal{E}) \leq 2^{-\eta}.$$

*Proof:* We first observe that

$$p_S(s) = \sum_{y: \varphi(y)=s} p_Y(y), \quad p_{Y|S}(y|s) = \frac{p_Y(y)}{p_S(s)}.$$

We have the following chain of inequalities:

$$\begin{aligned}
& p_{SXY}(\mathcal{D}^c \cap \mathcal{E}) \\
&= \sum_{s \in \mathcal{L}} p_S(s) \sum_{y: \varphi(y)=s} p_{Y|S}(y|s) \\
&\quad \times \sum_{\substack{x: \psi(s, \phi(x))=x \\ p_{X|S}(x|s) < (1/|\mathcal{M}|)2^{-\eta}}} p_{X|Y}(x|y) \\
&= \sum_{s \in \mathcal{L}} p_S(s) \sum_{\substack{x: \psi(s, \phi(x))=x \\ p_{X|S}(x|s) < (1/|\mathcal{M}|)2^{-\eta}}} p_{X|S}(x|s) \\
&\leq \sum_{s \in \mathcal{L}} p_S(s) \frac{1}{|\mathcal{M}|} 2^{-\eta} |\{x : \psi(s, \phi(x)) = x\}| \\
&\stackrel{(a)}{\leq} \sum_{s \in \mathcal{L}} p_S(s) \frac{1}{|\mathcal{M}|} 2^{-\eta} |\mathcal{M}| = 2^{-\eta}.
\end{aligned}$$

Step (a) follows from that the number of  $x \in \mathcal{X}$  correctly decoded does not exceed  $|\mathcal{M}|$ . ■

*Proof of Proposition 1:* By definition we have

$$p_{SXY}(\mathcal{D}) = p_{SX} \left\{ \log |\mathcal{M}| \geq \log \frac{1}{p_{X|S}(X|S)} - \eta \right\}.$$

Hence, it suffices to show

$$P_c(\phi, \varphi, \psi|p_{XY}) \leq p_{SXY}(\mathcal{D}) + 2^{-\eta}$$

to prove Proposition 1. By definition we have

$$P_c(\phi, \varphi, \psi|p_{XY}) = p_{SXY}(\mathcal{E}).$$

Then we have the following.

$$\begin{aligned}
& P_c(\phi, \varphi, \psi|p_{XY}) = p_{SXY}(\mathcal{E}) \\
&= p_{SXY}(\mathcal{D} \cap \mathcal{E}) + p_{SXY}(\mathcal{D}^c \cap \mathcal{E}) \\
&\leq p_{SXY}(\mathcal{D}) + p_{SXY}(\mathcal{D}^c \cap \mathcal{E}) \stackrel{(a)}{\leq} p_{SXY}(\mathcal{D}) + 2^{-\eta}.
\end{aligned}$$

Step (a) follows from Lemma 1. ■

*Proof of Theorem 1:* We have the following chain of inequalities:

$$\begin{aligned}
P_{c, \max}^{(1)}(p_{XY}) &\stackrel{(a)}{\leq} p_{SX} \left\{ 0 \geq \log \frac{1}{p_{X|S}(X|S)} - \eta \right\} + 2^{-\eta} \\
&= p_{SX} \left\{ \left[ \log \frac{1}{p_{X|S}(X|S)} \leq \eta \right] \right. \\
&\quad \left. \cap \left[ \log \frac{p_{X|S}(X|S)}{p_X(X)} < \nu \right] \right\} \\
&\quad + p_{SX} \left\{ \left[ \log \frac{1}{p_{X|S}(X|S)} \leq \eta \right] \right. \\
&\quad \left. \cap \left[ \log \frac{p_{X|S}(X|S)}{p_X(X)} \geq \nu \right] \right\} + 2^{-\eta} \\
&\leq p_X \left\{ \log \frac{1}{p_X(X)} < \eta + \nu \right\} \\
&\quad + p_{SX} \left\{ \log \frac{p_{X|S}(X|S)}{p_X(X)} \geq \nu \right\} + 2^{-\eta} \\
&\stackrel{(b)}{\leq} p_X \left\{ \log \frac{1}{p_X(X)} < \eta + \nu \right\} \\
&\quad + \frac{1}{\nu} E_{p_{SX}} \left[ \log \frac{p_{X|S}(X|S)}{p_X(X)} \right] + 2^{-\eta} \\
&= p_X \left\{ \log \frac{1}{p_X(X)} < \eta + \nu \right\} + \frac{1}{\nu} I(X; S) + 2^{-\eta}. \quad (6)
\end{aligned}$$

In step (a) we use Proposition 1 for  $|\mathcal{M}| = 1$ . Step (b) follows from the Markov's inequality. In (6), we choose  $\eta, \nu$  so that

$$\eta + \nu = \log \frac{1}{p_{\max}} = \min_{x \in \mathcal{X}} \log \frac{1}{p_X(x)}.$$

Since

$$\eta = -\nu + \log \frac{1}{p_{\max}} > 0,$$

$\nu$  must satisfy  $\nu \in (0, \log \frac{1}{p_{\max}})$ . For this choice of  $\eta, \nu$ , we have

$$p_X \left\{ \log \frac{1}{p_X(X)} < \eta + \nu \right\} = 0, 2^{-\eta} = 2^\nu (p_{\max}). \quad (7)$$

From (6) and (7), we have the bound (5) in Theorem 1. ■

## REFERENCES

- [1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Sys. Tech. Journal*, vol. 28, pp. 656-715, 1949.
- [2] A. D. Wyner, "The wire-tap channel," *Bell Sys. Tech. Journal*, vol. 54, pp. 1355-1387, 1975.
- [3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 339-348, 1978.
- [4] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3989-4001, Jun. 2011.
- [5] T.S. Han, H. Endo, and M. Sasaki, "Reliability and secrecy functions of the wiretap channel under cost constraint," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6819-6843, Nov. 2014.